Evaluation of Fitness Functions of GA Classification

Ming-Der Yang
Dept of Civil Engineering
National Chung Hsing University
250 Kuokuang Rd, Taichung, Taiwan
886-4-22854387
mdyang@nchu.edu.tw

Yeh-Feng Yang
Dept of Civil Engineering
National Chung Hsing University
250 Kuokuang Rd, Taichung, Taiwan
886-4-22854387
yf.yang@yahoo.com.tw

Yi-Ping Chen
Dept of Business Administration
DaYeh University
168, University Rd., Changhua, Taiwan
886-4-8511888
ypchen@mail.dyu.edu.tw

ABSTRACT

Genetic Algorithm (GA) classifier can automatically search a proper clustering number according to fitness evaluation instead of assignment by users. In this study, a GA classifier with various fitness functions is adopted to search the cluster centers and a suitable cluster number for digital images to overcome the disadvantages of the conventional unsupervised classifier. By employing a proper clustering index as fitness, a GA with length-variable chromosome can determine the most suitable number of clusters and the most proper cluster centers.

This paper evaluates three popular classification indexes, including DBI, FCMI, and PASI, as fitness functions in GA operation. The GA classifier is applied to SPOT-5 satellite image to verify its accuracy and robustness. The results show the GA classifier adopting FCMI having the best performance, followed by DBI and PASI, sequentially. Regarding to computation efficiency, the GA classification with DBI took much less computation time of the GA classifications with FCMI and PASI.

Categories and Subject Descriptors

I.5.2 **[Pattern Recognition]**: Design Methodology - *Classifier design and evaluation, Pattern analysis*

Keywords

Genetic Algorithm; Fitness function; Overall Accuracy; Unsupervised Classification.

1. INTRODUCTION

Genetic algorithm (GA) inspired by natural selection and survival of the fittest produces a serial processes simulating biological evolution to solve real-world problems. GA has been widely and successfully used in image processing, data interpretation, artificial intelligence, and many other areas. In particular, unsupervised image classification can be processed through GA operations, such as natural selection, cross-over, and mutation, to provide the optimal solutions. Multispectral SPOT imagery was adopted for overall affecting scoring by combining terrain factors through an aggregation function to produce a synthetic probability map of landslide reoccurrence for landslide reoccurrence in the Tsao-Ling area [1]. A GA-classifier was proven to be capable of performing automatically as well as a supervised classifier. Martini et al. employed multi-date images of Synthetic Aperture Radar to identify dry snow cover in Alps mountainous region in France [2]. DeAlwis et al. transformed Landsat 7 ETM+ satellite images into Normalized Difference Water Index (NDWI) that was processed through unsupervised classifier ISODATA for monitoring the storage capability of underground water [3]. A GA-classifier was adopted to search the cluster centers and a suitable cluster number for digital images to overcome the disadvantages of the conventional unsupervised classifier [4].

Clustering validity indices were developed to determine optimal clustering, such as separation index (SI), Daviers-Bouldin (DB) index, Xie-Beni (XB) index, Hubert's statistics, and Dunn's index (DI) in which DB index has both a statistical and geometric rational [5]. Regarding to fitness functions in GA classification, Bandyopadhyay and Maulik adopted DBI (Davies-Bouldin Index), DI (Dunn's index), FCMI (Fuzzy C-Means index), and CI (C-index) in GA classifiers for clustering problem solving [6]. Among those fitness functions, the GA classifier with DBI was found to perform best in satellite image classification. Also, Bandyopadhyay and Maulik integrated conventional unsupervised classifier, K-Means, and GA classifier to overcome the requirement of the initial cluster number in K-Means classification, and obtained better classification results than original individual classifiers [5]. PSI (Partition separation index) was proposed for clustering, and was compared with popular indexes, such as PCI (Partition Coefficient Index), PEI (Partition Entropy Index), FSVF (Fukuyama and Sugeno Validity Function), XBI (Xie and Beni Validity Function), and DBI (Davies and Bouldin validity function) for verifying its superiority [7].

2. METHODOLOGY

This paper adopts three popular classification indexes, including DBI (Davies-Bouldin Index), FCMI (Fuzzy c-Means Index), and PASI (Partition Separation Index, as fitness functions in GA operation for comparison in classification accuracy and computation time. The calculation of DBI, FCMI, and PASI can be checked in Yang's paper [9]. This paper experiments chromosome lengths of 6, 8, 12, and 20, and crossover rates of 0.2, 0.4, 0.6, and 0.8. Other GA operation setting includes real-number coding, population of 30, selection way of Roulette Wheel Selection, crossover way of two-points crossover, and mutation rate of 0.03. To verify the GA classifier, a SPOT 5 satellite image were processed under various condition setting.

3. RESULT ANALYSIS

Figure 1 shows the results varying with different chromosome lengths. The class numbers determined by PASI are inclined to beneath the number of landuse. DBI presents the unstable numbers increasingly with the increasing chromosome length. The clustering results by FCMI present the moderate number of landuse and the distributions of the related species. The overall accuracy and K-HAT of FCMI classification are the highest, followed by DBI and PASI. In the comparisons of computation

GECCO'14, July 12–16, 2014, Vancouver, BC, Canada.
ACM 978-1-4503-2662-9/14/07.

time between the three indices, DBI classification is the most computation-efficient, whereas PASI classification is the least. Figure 2 presents the classification images determined by the three indices varying with the different crossover rates, and shows no obvious impact of crossover rate on cluster numbers. Regarding to overall accuracy and K-HAT, high crossover rates can result in more accurate results for DBI and FCMI classification, whereas PASI classification shows an opposite relationship between crossover rate and overall accuracy and K-HAT. Furthermore, PASI and FCMI, which are based on the fuzzy logical modeling, spent more computation time than DBI and cost more as the increase of crossover rate. Furthermore, PASI and FCMI, which are based on the fuzzy logical modeling, spent more computation time than DBI and cost more as the increase of crossover rate.

4. CONCLUSION

According to the previous result analysis and discussion, the paper concludes the following points:

(1) Through various setting of chromosome and crossover rate, this paper shows the classifier with FCMI preformed the best, followed by DBI and PASI, sequentially.

(2) The GA classifications with DBI and FCMI have high accuracy as crossover rate increasing.

(3) Regarding to computation efficiency, the GA classification with DBI took much less computation time of the GA classifications with FCMI and PASI.

(4) The GA classification with DBI with the advantages of high accuracy and less computation is suitable for classification on large digital data, such as satellite images.

5. REFERENCES

[1] Yang, M.D., Yang, Y.F., and Hsu, S.C. 2004. Application of remotely sensed data to the assessment of terrain factors affecting the Tsao-Ling landslide. Canadian Journal of Remote Sensing. 30(4), 593-603.

[2] DeAlwis, D.A., Easton, Z.M., Dahlke, H.E., Philpot, W.D., and Steenhuis, T.S. 2007. Unsupervised classification of saturated areas using a time series of remotely sensed images. Hydrology and Earth System Sciences. Vol.11, 1609-1620.

[3] Martini, A., Ferro-Famil, L., Pottier, E., and Dedieu, J.P. 2005. Dry snow extent monitoring in strong topography conditions. Geoscience and Remote Sensing Symposium, IGARSS '05 Proceedings IEEE International. Vol.8, 5440-5443.

[4] Yang, M.D. 2007. A genetic algorithm (GA) based automated classifier for remote sensing imagery. Canadian Journal of Remote Sensing. 33(3), 593-603.

[5] Bandyopadhyay, S., and Maulik, U. 2002(a). An evolution technique based on K-Means algorithm for optimal clustering. Information Sciences. 146(2002), .221-237.

[6] Bandyopadhyay, S., and Maulik, U. 2002 (b). Genetic clustering for automatic evolution of clusters and application to image classification. IEEE pattern recognition. 35(2002), 1197-1208.

[7] Yang, M.S., and Wu, K.L. 2001. A new validity index for fuzzy clustering. IEEE International Fuzzy Systems Conference, 89-92.

[8] Sivanandam, S.N., and Deepa, S.N. 2008. Introduction to Genetic Algorithms, Springer. Berlin, 442p.

[9] Yang, M.D., Yang, Y.F., Su, T.C., and Huang, K.S. 2014. An efficient fitness function in genetic algorithm classifier for landuse recognition on satellite images, Scientific World Journal, doi:10.1155/2014/264512

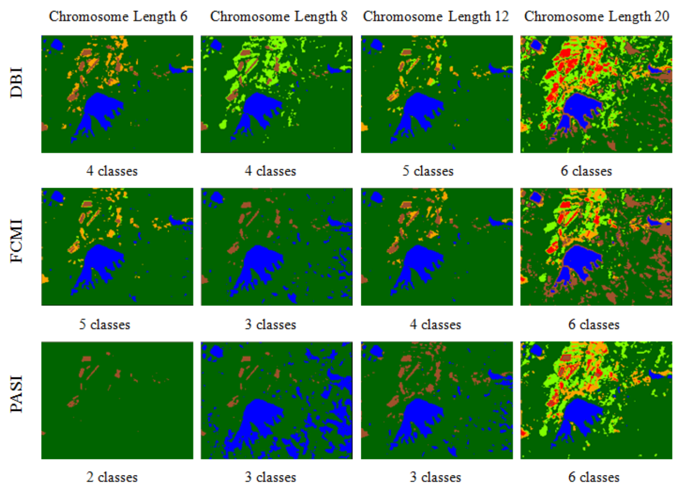

Figure 1. Classification varying with maximal centroid numbers

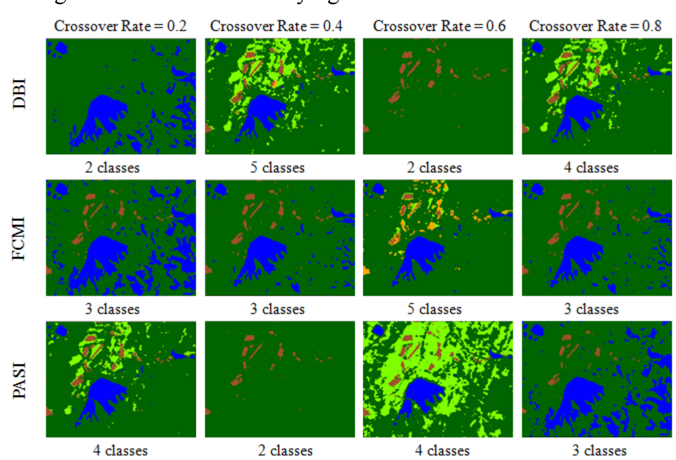

Figure 2. Classification determined by various populations